organized as a map 14 having a plurality of grids or units 16. Each unit 16 on the map 14 represents a collicular neuron that receives multisensory input from its corresponding location in the environment. The units 16, i.e., the model SC units 16 use sensory inputs such as video (V) 18 and audio (A) 20, for example, to compute the probability that something of interest, i.e., a target 22, has appeared in the surroundings.

Please replace the paragraph beginning on page 4, line 17, with the following rewritten paragraph:

The model 13 in accordance with one embodiment of the present invention approximates Bayes' rule for computing the probability of a target. Specifically, the SC units 16 in the map 14 approximate P(T|V,A), which is the conditional probability of a target (T) given visual (V) and auditory (A) sensory input. The Bayes' rule for computing the probability of a target given V and A is as follows:

Please replace the paragraph beginning on page 6, line 3, with the following rewritten paragraph:

The model 13 in accordance with another embodiment of the present invention estimates Bayes' rule for calculating target probability by using back-propagation which, as known in the art, is a supervised neural network learning algorithm. Generally, back-propagation is used to train neural networks having input units, output units, and units in between called hidden units. All units are sigmoidal. The input units send their activity to the hidden units, and the hidden units send their activity to the output units. The hidden and output units can also receive a bias input, which is an input that has activity 1 all the time.

All the connections between the input, output, and hidden units have weights associated with them. Back-propagation adjusts the values of the weights in order to achieve the desired output unit response for any input pattern. In the estimation method, the SC units 16 are the output units of neural networks that also have input and hidden units. The back-propagation algorithm is used to iteratively adjust the weights of the hidden and the output units to achieve the desired output.

Please replace the paragraph beginning on page 6, line 26, with the following rewritten paragraph:

Referring to FIG. 6, the acquisition phase includes acquiring raw video and audio input and preprocessing it (block 54), and applying the input to the neural network and finding the responses of the SC units 16 (block 56). Then, the SC unit 16 with the highest value is found (block 58). Using this information, the location corresponding to the SC unit 16 with the highest response value is chosen as the location of the next target (block 60).

Please replace the paragraph beginning on page 7, line 11, with the following rewritten paragraph:

Turning now to FIG. 7, the present unsupervised algorithm for approximating target probability includes two stages. The first stage involves an unsupervised learning mechanism that increases the amount of information transmitted from the sensory inputs, audio (A) and video (V), for example, to the SC unit 16 of the model SC 13. This mechanism is known in the art as the Kohonen mechanism, which has been shown to

increase information transmission in neural networks. The Kohonen mechanism is unsupervised, meaning that it would take the sensory inputs (such as audio and video) and automatically adjust the model SC 13 to increase the amount of information that is transmitted to it from the input. This is accomplished by adjusting the connection weights from the V and A inputs to the SC units 16 in such a way that individual SC units become specialized for specific inputs. For example, the Kohonen algorithm might cause one SC unit 16 to become specialized for video input from the extreme left side of the environment, and another to become specialized for audio input coming straight ahead. For very certain (not noisy) inputs, all the SC units 16 will become specialized for particular locations in the environment, and almost all of them will become specific for one modality or the other (V or A). The SC units 16 in this case can give a near maximal amount of information about the input. These units 16 can indicate not only where the target is but also of what modality it is.

Please replace the paragraph beginning on page 7, line 29, with the following rewritten paragraph:

If the inputs are not so certain (noisy), then the Kohonen algorithm will cause more of the SC units 16 to become bimodal and respond to both V and A. These SC units 16 would be less informative because they could indicate where the target is but not of which modality it is. Thus, the Kohonen algorithm will do the best it can with the input it is given to increase the amount of information that is transmitted to the SC units 16 from the V and A input units.

Please replace the paragraph beginning on page 8, line 20, with the following rewritten paragraph:

Cortical units 62 modulate the sensory inputs to the model SC units 16 by multiplying their weights. For example, the video input to an SC unit 16 would be $c_V w_V V$, where c_V is the amount of cortical modulation of that sensory weight w_V. In the learning process, an active cortical unit 62 will increase its modulation of a sensory input to an SC unit 16 if the SC unit is also active but the sensory input is inactive. If the SC unit 16 and the sensory input are both active then the cortical unit 62 will decrease its modulation of the sensory input. For example, when an SC unit 16 receives multisensory video and audio input after stage one training, and a target appears that provides a video input but produces no audio input, that SC unit will be active because it receives both video and audio input and the video input is active. A cortical unit 62 sensitive to video will also be active. Because the activity of the SC unit 16 and the cortical unit 62 are correlated, the cortical unit will change its level of modulation of the sensory inputs, accordingly as they are anti-correlated. Specifically, the cortical unit 62 will decrease its modulation of the video input (because the cortical unit and the video input are correlated) but increase its modulation of the auditory input (because the cortical unit and the audio input are anti-correlated).

Please replace the paragraph beginning on page 9, line 6, with the following rewritten paragraph:

Turning now to FIG. 8, the preferred embodiment for implementing the two-stage algorithm for approximating target probability involves iterative procedures that begin

after certain parameters in the model have been set. The structure of the neural network model 13 is determined in block 64, in which the number of SC unit 16 is set, and the bias weight and sensitivity of each SC unit are assigned. All the SC units 16 in the two-stage model are sigmoidal, where output y is related to input x by: $y=1/(1+\exp(-gx))$. The input x is the weighted sum of its inputs from V and A and from the bias. The bias weight w_b is the same fixed constant for all SC units 16. The sensitivity g is another fixed constant that is the same for all SC units 16. These fixed constants (w_b and g), along with the number of SC units 16, are set in block 64.

Please replace the paragraph beginning on page 10, line 12, with the following rewritten paragraph:

Referring now to FIG. 9, each iteration of the stage one learning process begins by acquiring and preprocessing the video and audio inputs from a randomly positioned target (block 76). These V and A inputs are sent to the SC units 16 over the ascending connections. As explained above, the sigmoidal SC units 16 use the weighted sum of these inputs to compute their responses (block 78). Then the SC unit 16 with the maximal response is found (block 80). The unit with the maximal response is referred to as the 'winning' SC unit. The ascending weights of the winning SC unit 16 and its neighbors are trained using Kohonen's rule (block 82). The neighbors of an SC unit 16 are simply the other SC units that are near it in the network. The number of neighbors trained in stage one is determined by the neighborhood-size parameter set in block 77 (see FIG. 8). Kohonen's rule basically adjusts the ascending weights to the winning SC unit 16 and its neighbors so that they become even more specialized for the current input.

Please replace the paragraph beginning on page 10, line 24, with the following rewritten paragraph:

Turning to FIG. 10, each iteration of stage-two learning process begins by acquiring and preprocessing the video and audio inputs from a randomly positioned target, and using that input to determine cortical activation (block 84). The term 'cortical' is meant to indicate that these units 62 are at a high level, as they are in the cortex of the mammalian brain, and the properties of the cortical units 62 can vary over a very broad range. For example, the cortical units can act as pattern recognizers, and can be specialized for particular types of targets like humans or airplanes. So far as applied here, the cortical units 62 simply register the modality of the target, whether it is visual, auditory, or both. A visual cortical unit 62, for example, would be active whenever the video input is active. Block 84 indicates that the activity of the cortical units 62 is dependent upon the video and audio inputs. The cortical units 62 send descending connections to the model SC units 16, and more specifically, to the connections onto the SC units from the V and A sensory inputs. As explained above, an active cortical unit 62 can modulate the weights of the ascending connections by multiplying the value of the ascending weight by that of the descending weight (block 86). After any cortical descending modulation of ascending weights is taken into account, the responses of the SC units 16 to the ascending input is computed (block 88).

Please replace the paragraph beginning on page 11, line 11, with the following rewritten paragraph:

Then the SC units 16 with responses less than cutoff are found and set to zero (block 90). Descending weights of SC units 16 are then trained using the following triple correlation rule (block 92):

> If an SC unit 16 and a cortical unit 62 are both active, then
> *increase* the descending weights to *inactive* ascending input synapses, and
> *decrease* the descending weights to *active* ascending input synapses.

Please replace the paragraph beginning on page 13, line 5, with the following rewritten paragraph:

The selection process is implemented by choosing the SC unit 16 that has the largest response to its inputs. Since the SC units 16 are in spatial register with their inputs, localization of the target is determined by the location of the chosen SC unit in the 1-dimensional array. Acquisition of the target then takes place by moving the rotating platform 102 to the coordinate in the environment corresponding to the chosen SC unit, thereby allowing the target 100 to be viewed by the operator through a monitor 116.

Please replace the paragraph beginning on page 13, line 11, with the following rewritten paragraph:

If target probability is obtained by estimating Bayes' rule using back-propagation, an array of computer-controlled buzzer/flasher pairs (not shown), spaced every 15 degrees, for example, is used to provide the sensory stimuli for back-propagation training. At each training cycle, one location is chosen at random, and the buzzer and the flasher at that location are activated. The 60 preprocessed audio and video signals are temporally summed or averaged over a window of 1 second and applied as input to the model. The inputs are applied to the model SC units 16 directly, or indirectly through a network of hidden units. The location of the source is specified as a 60 element desired output vector of 59 zeros, and a one at the location in the vector corresponding to the location of the source. The weights are all trained with one cycle of back-propagation, and the process is repeated with a source at a newly chosen, random location.

B1^2

B1^3

Please replace the paragraph beginning on page 13, line 22, with the following rewritten paragraph:

After training, the inputs are preprocessed as described above over the 1-second window and then applied in spatial register to the SC model 13. Each SC unit 16 then estimates, on the basis of its video and audio inputs, the Bayesian probability that the source is present at its corresponding location in the environment, which simulates MSE. The location of the model SC unit 16 with the largest response is then chosen as the location of the most probable target, and the camera 96 and the microphone 98 are aimed in that direction. The SAC system 94 chooses as targets those objects in the environment that move and make noise, which covers most of the targets actually chosen by the SC in guiding saccadic eye movements in animals.